

Enabling time travel for the scholarly web

July 16, 2013



Banishing the dreaded Internet search where 30 percent of research paper hyperlinks fail to connect

LOS ALAMOS, N.M., July 16, 2013—An international team of information scientists has begun a two-year study to investigate how web links in scientific and other academic articles fail to lead to the resources being referenced.

This is the focus of the Hiberlink project in which the team from Los Alamos National Laboratory and the University of Edinburgh will assess the extent of “reference rot” using a vast corpus of online scholarly work. It is funded by a grant of \$500,000 (£310,000) from the US-based Andrew W. Mellon Foundation, coordinated by EDINA, the designated online services center at the University of Edinburgh, which serves the needs of universities and colleges across the UK.

“Increasingly, scientific papers contain links to web pages containing, for example, project descriptions, demonstrations, and software. But, as we all know, web pages change or disappear,” said Herbert Van de Sompel, the Los Alamos principal investigator on the project. “Currently, there is no archival infrastructure to safeguard such pages and hence revisiting them some time after they were linked from a paper is many times impossible. The result is a broken scholarly record.”

Increasingly, web-based scholarship includes links that point to resources needed or created in research activity, including software, datasets, websites, presentations, blogs, videos etc. as well as scientific workflows and ontologies. These referenced resources often evolve over time, unlike traditional scholarly articles. The reference-rot problem occurs whenever the original version of a linked resource is not available anymore.

The problem has two aspects. First, the http:// link that references a resource may no longer function. Second, the content at the end of the link may have evolved and may even have become dramatically different from when originally referenced. So when eventually a researcher revisits an online scholarly work and double-checks referenced resources to confirm evidence or establish context, the original online information may have changed or even ceased to exist.

The Hiberlink project builds directly upon a pilot study from Los Alamos, powered by their Memento “Time Travel for the Web” technology that confirmed that as much as 30 percent of the http:// links in a selection of 400,000 arXiv.org papers did not function and that 65 percent of the remaining links referred to a resource that was not archived, and hence in danger of disappearing without a trace.

Using the text mining and information extracting tools by the Language Technology Group (LTG) at the University of Edinburgh School of Informatics, the project will examine a vast body of scholarly publications in order to assess which links still work as intended and what web content has been successfully archived and therefore preserved for use by future researchers and students.

The ultimate goal for the Hiberlink project is to identify practical solutions to the reference-rot problem, and to develop approaches that can be integrated easily in the publication process. The project leaders plan to work with academic publishers and other web-based publication venues to ensure more effective preservation of web-based resources so to increase the prospect of continued access for future generations of researchers, students and their teachers.

Links for more information:

- [EDINA](#)
- [Language Technology Group, Informatics at the University of Edinburgh](#)
- [Los Alamos National Laboratory, Research Library](#)
- [The University of Edinburgh](#)
- [Los Alamos National Laboratory](#)
- [Memento](#)
- [Pilot study paper](#)

Photo caption: From left, Martin Klein, Robert Sanderson and Herbert Van de Sompel, information scientists at the Los Alamos National Laboratory Research Library, discuss the information pathway involved in preventing “reference rot” in scientific material linked to the web.

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Operated by Los Alamos National Security, LLC for the Department of Energy's NNSA

